EXHIBIT 3

MANUSCRIPT TITLE

Implications of a MTBE Consumer Threshold Odor Study for Drinking Water Standard Setting

AUTHORS AND AFFILIATIONS

Andrew J. Stocking[1], Irwin H. Suffet[2], Michael J. McGuire[3] & Michael C. Kavanaugh[4*]

[1]Project Engineer
Malcolm Pirnie, Inc.
180 Grand Ave. Suite 1000
Oakland, California 94612-3754

[2] Professor
Environmental Science and Engineering Program
UCLA, School of Public Health, Room 46-081, CHS
Charles E. Young Drive South
Box 951772
Los Angeles CA 90095-1772

[3]President
McGuire Environmental Consultants, Inc.
1919 Santa Monica Blvd., Suite 350
Santa Monica, CA 90404-1950

[4]Vice President
Malcolm Pirnie, Inc.
180 Grand Ave. Suite 1000
Oakland, California 94612-3754

[*] To whom correspondence should be addressed
ph (510) 451-8900
fax (510) 451-8904
mkavanaugh@pirnie.com

Submitted
June 12, 2000

Implications of a MTBE Consumer Threshold Odor Study for Drinking Water Standard Setting

Andrew J. Stocking, Irwin H. Suffet, Michael J. McGuire, Michael C. Kavanaugh

Abstract

In late 2000, the USEPA is expected to adopt a secondary maximum contaminant level (SMCL) for MTBE in drinking water. This paper presents for the first time a consumer study to determine the taste and/or odor threshold of MTBE in drinking water. A protocol was developed based on ASTM method E679-91 and modified to address concerns raised by interested stakeholders. The study conducted according to the final odor threshold protocol, yielded a 15 ug/L odor consumer threshold for MTBE in drinking water using a panel of 57 consumers. The 15 ug/L is the geometric mean of the individual thresholds for each of the 57 consumers. This consumer panel threshold is consistent with the trained panel thresholds reported from five other taste and/or odor studies, which ranged from 13.5 to 45.5 ug/L. Consequently, the authors of this report recommend using the methodology presented in this paper as the basis for establishing the federal SMCL for MTBE and other organic chemicals in drinking water.

Key words: MTBE, odor, odor threshold concentration, drinking water standard,

Introduction

Methyl tert-butyl ether (MTBE) has been used as a gasoline additive in increasing volume percentages since 1979. In 1998, it was added to approximately 30% of the nation's gasoline at approximately 11-15% by volume[1]. As a result of this widespread usage and its unique physio-chemical properties, MTBE has been detected in drinking water wells in areas with leaking underground storage tanks[2,3]. Consequently, setting protective drinking water standards for MTBE has become an issue for regulatory agencies in federal and state governments. As of June 2000, the US Environmental Protection Agency (USEPA) had not yet set any drinking water standards for MTBE. They had, however, issued a health advisory in

December of 1997 to provide guidance to communities whose water supplies had been impacted by MTBE. The advisory recommended that "keeping the concentrations [of MTBE] in the range of 20 to 40 ug/L or below will likely avert unpleasant taste and odor effects"[4], thus suggesting that MTBE is considered more of a taste and odor issue than a health issue. The USEPA is expected to propose and implement a formal SMCL for MTBE in late 2000. This represents the first time that the USEPA will develop a secondary MCL based upon taste and odor for a specific organic chemical.

At the state government level, several states, including Maine, California, New Hampshire, and New Jersey, have set primary maximum contaminant levels (MCLs) for MTBE. Maine and New Jersey have set MTBE MCLs of 35 and 70 ug/L, respectively, while in May 2000, California and New Hampshire adopted MCLs of 13 ug/L. All of these standards are based on protection of human health and not aesthetic concerns. California is unique in setting a secondary MCL (SMCL) for MTBE based on the organoleptic characteristics of MTBE in water; it was adopted on January 7, 1999 at 5 ug/L. In contrast to all other states, SMCLs in California are enforceable standards. The technical foundation for the California SMCL for MTBE was derived from 1) an estimation of the first concentration that could not be detected by any of the Young et al.[5] testing panel, and 2) the lowest concentration detected by any panelist in the Shen et al.[6] study that would not be subject to background laboratory MTBE interference during analytical measurement. This SMCL represented California's second attempt to establish an SMCL for an individual organic compound. However, as the final statement of reasons for the MTBE SMCL illustrates, disagreements exist over the interpretation and application of these previous studies to drinking water standard setting[7].

Scope and objective

Since 1996, five independent studies have been completed to determine the taste and odor thresholds for MTBE in drinking water[5,6,8,9,10]. Each of these studies used trained expert panelists to elicit the threshold value for MTBE in drinking water. In light of the regulatory developments

in California and elsewhere in early 1998, representatives of the regulatory and industrial communities believed that additional taste and/or odor data for MTBE should be developed using an untrained consumer panel, which was believed to reflect more accurately public sensitivities. Therefore, the Oxygenated Fuels Association (OFA) commissioned Malcolm Pirnie to organize a consumer taste and/or odor study with a large sample population to augment the existing five taste and odor studies. Of critical importance in this process was the development of a methodology that would serve as a solid foundation for collecting data that could be used to support an SMCL, and serve as a precedent for establishing future SMCLs for other organic chemicals.

This paper presents the methodology, results, and a discussion of the implications of these results for setting a SMCL for MTBE. The results and conclusions of this paper are intended to be generally applicable; however, much of the discussion addresses the California experience in establishing SMCLs.

SMCL regulatory authority and guidance

In accordance with the Safe Drinking Water Act (SDWA) of 1974, as amended in 1986 and 1996, the US Congress directed the USEPA to develop and enforce primary, health-risk based standards, or maximum contaminant levels (MCLs)[11]. Under the SDWA, each state is given the option to 1) adopt the federal standards or develop more stringent standards, and 2) allow the federal government to retain enforcement responsibility or take on enforcement responsibility from the federal government. Most states have chosen to locally enforce the federal standards.

In addition to primary MCLs, the SDWA required that the federal government establish secondary, aesthetic-based, maximum contaminant levels (SMCLs) for organic and inorganic chemicals in water. Again, most states have chosen to adopt the federal secondary standards; however, due to the aesthetic-based nature of these standards, they are generally not enforced by the federal or state governments. Thus, states neither require frequent monitoring for these

chemicals nor do they levy fines or take legal actions when secondary standards are exceeded. In contrast, the California statute addressing drinking water declares: "It is the intent of the Legislature to improve laws governing drinking water quality, to improve upon the minimum requirements of the federal Safe Drinking Water Act Amendments of 1996, . . ."[12]. Thus, California is one of the few states to establish several primary standards more stringent than corresponding federal standards and to enforce all secondary standards.

The SDWA provides several rules that must be followed when establishing primary and secondary drinking water MCLs. Primary MCLs are based on avoiding exposure of the population to individual chemicals based on chemical-specific health risk levels. The bases for setting these standards must conform to several sections of the SDWA, namely, *Use of Science in Decision Making, Health Risk Reduction and Cost Analysis, Feasibility, and Subpopulations at Greater Risk*[11]. The collective intent of these sections is to ensure that primary standards are scientifically supported, protective of the population, and technically and financially feasible.

However, the development and implementation of SMCLs is only vaguely described in federal regulatory language. Under the SDWA, a secondary drinking water standard is applied to any contaminant in drinking water which:

A) may adversely affect the odor or appearance of such water and consequently may cause a substantial number of the persons served by the public water system providing such water to discontinue its use, or

B) may otherwise adversely affect the public welfare.

The SDWA follows by noting that, "Such regulations may vary according to geographic and other circumstances"[11]. Thus, due to the lack of specific regulatory guidance for setting SMCLs, aside from the above excerpts, states have significant latitude when establishing SMCLs.

Based on this guidance, the USEPA has set many SMCLs for a variety of taste, odor, and color causing constituents in drinking water (see Table 1) since 1974. These regulated constituents are primarily inorganic chemicals, such as aluminum, iron, magnesium, or general characteristics of water, such as color, corrosivity, odor, but do not currently include an individual organic chemicals. Currently, taste and odor in water from organic and inorganic sources is regulated federally and by states using the Threshold Odor Number (TON), which is designed to represent the minimum number of sample dilutions required to achieve a detectable taste or odor in water. By definition, a TON of 1 indicates that if a sample is diluted at all, the sample has no detectable odor to the trained assessor[13]. If a sample is assigned a TON of 3, for example, it indicates that the sample must be diluted to three times its initial volume to reach the lowest detectable concentration. However, the TON method has been documented to be unable to guarantee an aesthetically acceptable water quality in many situations, especially when earthy, musty odors are present[14]. Thus, other methods including the Flavor Profile Analysis (FPA) were developed to protect consumers from taste and odor problems[15].

Background of taste and odor testing procedures

Taste and odor (T&O) evaluation and quantification is continuing to evolve as an applied science[16]. The American Society for Testing and Materials (ASTM) and American Public Health Association (APHA) have developed procedures and practices that specify standardized applications of sensory methods. The methods use panelists as measuring devices, which is analogous to the use of analytical instruments to quantify the concentration of specific chemicals. Taste and odor threshold concentrations for specific compounds in water are set using a taste and odor threshold test as outlined in Standard Method 207 for odor and Standard Method 211 for taste[13], as well as ASTM E679 for odor[17].

The development of a taste and odor testing panel is fundamental to each of these methods due to differences in individual sensitivity to T&O causing compounds. There are two types of panels for determining taste and odor thresholds: expert panels and consumer panels.

Expert, or trained panelists are people with increased odor sensitivity. While this may be desirable for regular monitoring of odorous substances in drinking water before the water is treated or distributed, it may skew the results of an odor threshold determination study toward the more sensitive portion of the population. However, trained panelists are capable of characterizing and describing odors and will, therefore, not be expected to guess whether a taste or odor is present.

A consumer (i.e., untrained) panelist is capable of determining whether a difference is present between multiple samples and is often required to guess which sample is different. For example, in a forced choice triangle test, such as ASTM E679-91, a consumer panelist is forced to identify the *different* sample, even if a taste or odor is not detectable. The risks of false positives are thus higher for a consumer panel compared to a trained panel. The consumer panel test is designed to determine the detection threshold, or the threshold where a consumer can detect a difference, but not identify or characterize that difference. Regardless of the type of panel, it is important that panelists be diverse in terms of age, gender, and ethnicity, as each of these factors can affect sensory perceptions.

Most odor studies use trained, expert panelists because of their familiarity with taste and odor testing protocol and the need for fewer panelists to calculate a threshold. Standard Methods states that, "when the results must represent the population as a whole or when great precision is desired [it is recommended that expert panels contain] not less than five persons, and preferably ten or more"[18]. As this statement implies, statistical analysis of the results from T&O studies is a complex issue, and particularly dependent on the panel size and composition. Typically, threshold standards for various chemicals are determined using the geometric mean of the T&O thresholds from panelists' results, in which case, a panel of 8-10 trained panelists is usually sufficient[13,18].

However, if an untrained consumer panel is used, additional panelists are needed to achieve scientifically valid results due to the likelihood of much larger variability in each consumer's sensitivity[18]. As more panelists are added, individual panelist's results become less

important and there is a larger numeric confidence that the panelists' results are representative of some larger population under similar testing conditions. To statistically represent an even larger population, such as the state of California, with any significant level of confidence is even more difficult. Often studies attempt to be protective of a larger percentage of the population by choosing a panel that is more sensitive to the chemical of concern. The geometric mean of this panel gives a conservative estimate of the population threshold under similar testing conditions. If the panel consists of the most sensitive individuals from the population, then the geometric mean gives an estimate of the minimum concentration detectable to the population.

With this policy and technical background for setting SMCLs, the remainder of this paper will focus specifically on the odor threshold study completed for MTBE and policy implications of that study.

Previous taste and/or odor studies for MTBE

At least five taste and/or odor studies for MTBE in water have been completed[5,6,8,9,10]. Each of these studies relied on published taste and odor standard methods as a basis for determining threshold detection values for MTBE (see Table 2). Each of these studies used expert panelists to describe the odor associated with MTBE in water. The most frequently used descriptors to characterize the odor of MTBE were "sweet solvent"[10] and "estery, vanilla, and sweet"[5]. However, at concentrations greater than 20 ug/L, the descriptor "solvent" was used more frequently and at an increasingly greater intensity[10]. The API and ARCO studies determined both the detection (39 ug/L to 149 ug/L) and recognition (44 ug/L to 212 ug/L) threshold for MTBE[8,9]. The significance of the difference between detection and recognition thresholds is that based on regulatory experience, consumers will not complain about the quality of their drinking water until the recognition threshold is exceeded. The MTBE concentration is significantly higher for the recognition threshold as compared to the detection threshold based on these two studies.

In cases where the published method failed to provide specific guidance on a procedure, each of the five studies made procedural assumptions, such as the type of water to use and the method of comparison. Similarly, several of the studies deviated from the published procedures for analyzing data. ASTM Method E679-91 states that the "best-estimate" threshold for any individual consumer is the geometric mean between the last miss and the first detection. Detection is defined as the concentration at which the consumer successfully detects all higher concentrations. Although data analysis techniques other than that described by ASTM E679-91 may be more rigorous when analyzing taste and odor threshold data, these methods should be justified and verified prior to use.

However, both the ARCO and API studies presented results based on an alternative numerical analysis of the data—an extrapolation of the data based on a linear regression of the data, assuming a log normal distribution of the results. Using this alternative analytical technique, the thresholds from both of these studies resulted in detection thresholds higher than the Orange County study, Young study, or MWD study. As a result of these deviations from published methods or from similar tests, the applicability of the ARCO and API results for setting SMCLs is questionable. To overcome this deficiency, a retest at concentrations encompassing the reported odor threshold concentrations should be completed to confirm the results of the two studies.

Scope of consumer panel study

To augment results from previous studies and to avoid the deficiencies in the other tests as noted, the current study was undertaken[19]. Malcolm Pirnie retained the National Food Laboratory (NFL) to organize the consumer panel and execute the sensory evaluation test. This consumer study was designed to be complimentary to previous studies, and to correct the main limitations of those studies, namely, the size and makeup of the panel. Considerable efforts were expended to incorporate the experiences of the previous studies through discussions with some of the principal investigators. Following compilation of a draft protocol, Malcolm Pirnie worked

closely with members of the Association of California Water Agencies (ACWA) to receive recommendations for improvement.

To resolve inconsistencies among commentors, the stakeholders agreed to the formation of an Expert Advisory Panel consisting of Dr. Mel Suffet and Dr. Michael J. McGuire—two past chairmen of the AWWA Taste and Odor Committee—and Dr. Richard Berk of the University of California, Los Angeles—a statistician familiar with taste and odor study statistical analyses. Together, this panel incorporated all comments from ACWA in addition to verbal comments received from Dr. Steve Book of the California Department of Health Services to finalize the protocol. Ultimately, the study protocol, including the procedure for statistical data analysis, followed ASTM Method E679-91 primarily because this methodology is standard in the food industry and could be rigorously duplicated by other researchers.

Protocol development

The final protocol was based on a consensus of the authors and Dr. Berk and deemed responsive to the ACWA comments. An odor study was chosen in lieu of a taste (flavor) study for two reasons: 1) an odor study was thought to result in a lower threshold[5]; and 2) there were no laboratories available that would accept the liability of performing a taste (flavor) study without a primary maximum contaminant level established by the California DHS.

The ASTM method describes a forced choice triangle test where the consumer must choose one of three samples as different from the other two. Although this inherently requires panelist guessing, the triangle method was determined to be the best published method available. While the ASTM method describes many of the required details for performing an odor threshold study, there are a few aspects that the method does not describe, including:

- maximum number of trials to present to a consumer,
- type of water to use for the blanks,
- sample presentation container, and
- water temperature.

These details were addressed by the authors in cooperation with Dr. Berk and a consensus was reached on each issue. The odor protocol used for this study involves presenting eight samples to the panelists in increasing concentrations (see Table 3), distributed according to a log-normal distribution. The MTBE and blank comparison solutions were prepared in odor-free bottled water at room temperature. The authors agreed that eight trials were the maximum number of trials a consumer could be exposed to before olfactory fatigue began to impact the results[18]. The authors also agreed based on FPA analysis that bottled water with less than 500 mg/L TDS, as opposed to deionized or distilled water, was a neutral water that would neither mask nor enhance any MTBE odor in the samples. Room temperature was chosen by the authors as the operating temperature because it allowed this study to be comparable to other studies and it represented a temperature at which the public most often consumes drinking water.

The samples were presented to the panelists in disposable plastic cups that were determined by the National Food Laboratory (NFL) to be odor-free. Plastic cups were used instead of glass containers because glass containers often retained residual odors. Each spiked and blank sample was covered with a clean watch glass. The panelists were instructed to lift each sample, swirl it several times, lift the watch glass, and smell the sample, as suggested in the FPA protocol[15]. The panelists were allowed to repeat the trial if they were uncertain after the first time. Once completed, the panelists replaced the watch glasses and signaled to the NFL staff that they were finished. The panelist then marked on a scorecard the number of the sample that smelled different from the other two. If the panelist was not able to determine a difference, they were told to guess which sample smelled different.

A key aspect of the protocol development was the composition of the panel. The original intention of this study was to use a large (50+) consumer panel. While a random-digit dialed consumer panel recruited from across California may have been a stronger statistical representation of the California population, the authors and Dr. Berk acknowledged that this was not possible given the available budget, time frame, and logistical constraints of the study. Therefore, the test panel was recruited from a database of over 10,000 consumers available from

the NFL. Given the constraining circumstances, the authors and Dr. Berk decided that this database of consumers was sufficient to develop a consumer panel odor threshold.

Many of the consumers from the NFL database had participated in past NFL sensory testing, ranging from beer tasting to ice cream evaluations. NFL stated that only a few consumers in the database had done any type of drinking water evaluation in the past. The authors and Dr. Berk agreed that consumers recruited from this database should represent a cross section of ages and genders who have not participated in sensory testing for at least one year prior to this study. NFL recruited the panelists according to these guidelines (see Table 4). Consumers who smoked, were pregnant, or had been diagnosed as asthmatic were not used because NFL wanted to limit potential auxiliary odors and liability concerns.

Once at the NFL testing area, the consumers read and signed a disclosure statement and received a brief orientation describing the testing process. Before testing began, NFL conducted an example triangle test involving touching different grades of sandpaper to familiarize the panelists with the triangle test methodology. However, the consumers were not familiarized with the odor of MTBE, because the authors felt that this would "train" the consumer panel, which is not desired for consumer testing.

Due to size limitations of the NFL testing facility, five testing sessions over two days were conducted with 9-14 consumers at each session. Between sessions, NFL used fans to dissipate any fugitive MTBE odors present in the testing areas. NFL prepared fresh MTBE spiked samples for each test session and prepared two split solutions for the subsequent Quality Assurance/Quality Control (QA/QC) validation of the test concentrations. Gas chromatography/mass spectroscopy (GC/MS), EPA Drinking Water Method 524.2, was performed on all of the samples for each consumer panel. The GC/MS analysis results fell within the required range: ±40% for values < 10 ug/L and ±20% for values > 10 ug/L (see Table 5), except for the few deviations noted.

Results